(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 October 2003 (23.10.2003)

PCT

(10) International Publication Number
WO 03/088132 A1

(51) International Patent Classification[7]: G06K 9/00, 9/32, 9/46

(21) International Application Number: PCT/SG02/00060

(22) International Filing Date: 12 April 2002 (12.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: KENT RIDGE DIGITAL LABS [SG/SG]; 21 Heng Mui Keng Terrace, Singapore 119613 (SG).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: MARIANI, Roberto [FR/SG]; Spanish Village, Farrer Road, Block 56 #01-05, Singapore 246688 (SG).

(74) Agent: AXIS INTELLECTUAL CAPITAL PTE LTD; 21A Duxton Road, Singapore 089487 (SG).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

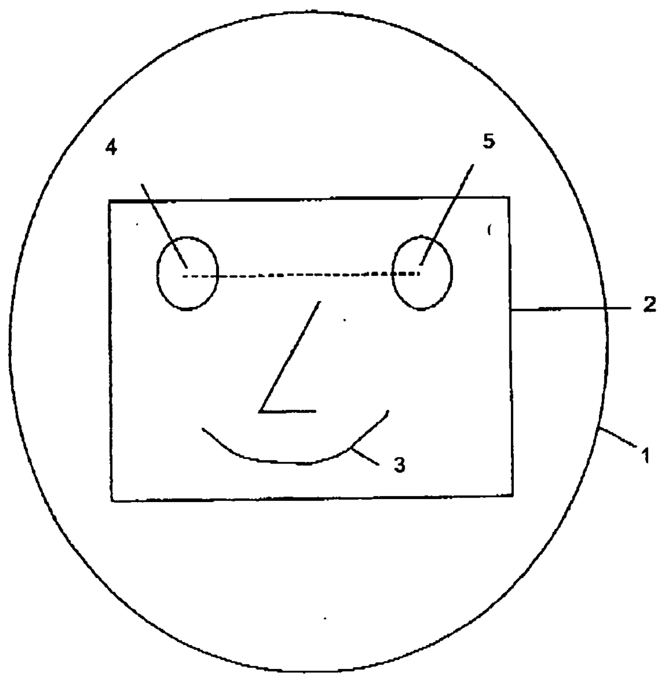
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ROBUST FACE REGISTRATION VIA MULTIPLE FACE PROTOTYPES SYNTHESIS

(57) Abstract: A face recognition and/or verification system including the step of registering a persons actual face wherein an image of the actual face is captured and synthesized to create a plurality of face prototypes, and wherein the face prototypes are stored for later analysis and comparison with a captured image to be recognised or verified.



WO 03/088132 A1

5/PRTS
DT04 Rec'd PCT/PTO 12 OCT 2004

ROBUST FACE REGISTRATION VIA MULTIPLE FACE PROTOTYPES SYNTHESIS

FIELD OF THE INVENTION

The present invention is generally directed towards face recognition and face verification systems and in particular towards a facial prototype synthesis system able to generate realistic templates and provides a more robust system for registering a face.

BACKGROUND OF THE INVENTION

The classical face recognition approach is to store an image of a person's face, and then to provide a face matching algorithm robust enough to handle varying lighting conditions, facial expressions, face directions, glasses, beard, mustache and facial hair, etc.

In the area of face recognition technology, research has focused almost entirely on developing algorithms that are invariant to the lighting conditions, the facial expressions and the face direction. Such systems obtain simple databases at the expense of complex matching algorithms or family of algorithms. Alternatively, the face recognition systems based on face template matching and neural networks, require a large number of face samples to train the network to an acceptable standard. The operations that are applied to train neural networks are mainly of linear geometric nature, such as scaling or zooming.

The problems of these techniques is their weakness in dealing with various lighting conditions, changes of expression as well as time difference between the registration and the time of the observation.

For example, WO 99/53427 provides a technique for detecting and recognizing an object in an image frame. The object identification and recognition process uses an image processing technique based on model graphs and bunch

SUMMARY OF THE INVENTION

With the above objects in mind the present invention provides in one aspect a face recognition and/or verification system including the step of registering a persons actual face wherein an image of the actual face is captured and synthesized to create a plurality of face prototypes, and wherein the face prototypes are stored for later analysis and comparison with a captured image to be recognised or verified.

The face prototypes may represent possible appearances of the actual face under various lighting conditions, varying facial expressions, varying facial orientations, and/or modeling errors.

In a further aspect the present invention provides a facial prototype synthesis system wherein an image of a persons actual face is synthesized to create a plurality of face prototypes, said face prototypes representing possible appearances of said actual face under various lighting conditions, varying facial expressions, varying facial orientations, and/or modeling errors, and wherein said face prototypes are stored for later use.

In the preferred embodiment the system normalises the captured image such that the eyes are a fixed distance apart and on a horizontal plane. The system may only synthesize the area of the face bounded by the eyebrows and mouth on the assumption that other features such as hair do not significantly alter for recognition and verification purposes. The synthesis may include consideration of alternate eye positions and applying masks to account for changing conditiors and errors.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the area of the face analysed in the preferred embodiment.

Figure 2 shows possible eye positions.

Figure 3 shows the use of geometric masks.

Figure 4 shows the use of optical flow approximations.

Figure 5 shows the use of exponential and logarithmic functions.

Figure 6 shows an example function for use with a vertical shadow mask.

Figure 7 shows the use of shadow filters.

Figure 8 shows an example of manual lighting masks.

Figure 9 shows an example of registering a user in accordance with the preferred embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENT

As it can be difficult to achieve accurate face recognition due to factors such as changing lighting conditions, the present invention proposes a method or system which relies on a more intensive registration process.

Most existing systems will capture an image of a person, and store that image for later comparison. In the present invention, during the registration of a person's face, the system automatically synthesizes a multitude of face prototypes, by creating artificial lighting conditions, artificial face morphing and by modeling the errors of a face location system, especially in the eyes detection process. These face prototypes represent the possible appearances of the initial face under various lighting conditions, various expressions and various face orientations, and under various errors of the face location system. The system obtains for each face, a set of faces that spans the possible appearances the face may have.

Having generated this multitude of face prototypes, data analysis can be applied, like dimensionality reduction (principal components analysis), feature extraction, automatic clustering, self-organising map. The design of a face recognition system based on these face prototypes can also be achieved. Face recognition systems based on face templates (pca) and/or feature vectors (gabor wavelet) may be applied, and they may also use these face clusters for training.

Considering now an example and referring to the figures. Given an image of a person's face and given the position of the two eyes (4, 5), the preferred system first normalizes the image by setting the two eyes (4, 5) on a horizontal plane and at a fixed distance apart, for example 50 pixels. This operation can be realized by a change of scale and rotation. Once the normalized image is created, the system selects the part of the face that is very stable across time, that is the system does not consider the part of the face above the eyebrows and below the mouth 3. That is, the system considers only the part of the face 1 within the box 2 shown in figure 1. Alternatively, other face location systems, may provide the 'face center', from which an estimate of the eyes position can be derived according to human anthropomorphic measures.

Using an existing face location system, it is likely that the eyes' position are imprecise. To propose a robust face encoding system, the present system assumes an error position of the eyes, and for each couple of eyes' positions, the system crops the face accordingly. In practice, the system preferably uses five possible positions per eye, leading to 25 cropped images, as illustrated in Figure 2. Of course, this number can be changed. By nature, this technique encodes rotation, translation and change of scales, as the scale factor is affected by the distance between the two eyes, and the rotation factor is affected by the angle of the two eyes with the horizontal line. In Figure 2, the dots 6 are the eyes detected by the system and the dots 7 are the other possible eyes' positions considered for the registration.

The circle 8 shows the surface error that describes the probable real position of the eye. Any point in this circle could be the real eye center. The ray of the surface error is fixed automatically by the system and depends on the interocular distance between the calculated position of both eyes. For example, for 50 pixels eyes distance, we may say that the eye position is precise at -/+5 pixels.

On each of these cropped and normalized images, the system applies predefined 2D lighting masks and predefined 2D warping or morphing masks and obtains a multitude of face prototypes that are likely to appear during the face location. The system preferably uses three families of lighting masks and 25 families of warping masks although varying numbers can be used. Any number of lighting and warping masks could be used, for example desirable results are being obtained using 16 lighting masks.

The system has two kinds of operations that are applied to an input image of a person in order to obtain an output image for storage. Namely, the photometric modification and/or the geometric modification of the image. The photometric modification changes the grey level or the color of a pixel, and the geometric modification modifies the position of this pixel.

The photometric transform can be encoded into the lighting masks and the geometric transform can be encoded into the morphing masks.

The geometric masks can be estimated by various ways, such as manually, semi-automatic or by optical flow estimation between two images. The optical flow estimation is a fully automatic technique exemplified in Figures 3 and 4.

The first row of Figure 3 has three original images, and the last row shows the three generated frontal face images obtained using geometric masks. Here the masks are tuned to generate frontal faces.

In the same way, the first row of Figure 4 contains five input images, the second row contains an approximation of the optical flow between each face and its vertical mirror, and the last row contains an approximation of the frontal face, good enough for a robust face recognition.

Here, we describe five photometric masks. Any number can be generated, but in testing these have proved to be very good for robust face registration, as they approximate real lighting conditions.

These preferred marks are:

i) Logarithmic function on grey-level; which obtains brighter images;

ii) Exponential stretch on the function; which obtains darker images;

iii) Vertical shadow that creates vertical half-shadowed faces;

iv) Horizontal shadow that creates horizontal half-shadowed faces; and

v) By differentiating images captured from the camera during the masks settings.

If we consider an input image with grey levels ranging from 0 to 1 after a standard grey level normalization process of the form (v-vmin)/(vmax-vmin), where vmin and vmax are respectively the minimum and maximum grey level values of the image, then each of the preferred masks may be described as follows:

Logarithmic function

Let [Kmin, 255] such that 255 > Kmin > 0. The system builds the lookup table that contains 256-Kmin entries by computing:

LOGLUT[w] = (log(w) + Kmin) – log(Kmin)) / (log(255)-log(Kmin)),

For w = Kmin,...255. The lookup table values are all ranging from 0 to 1. Given the value v of a pixel ranging between 0 and 1, the system

obtains the new grey level value w by w = LOGLUT[(255-Kmin)*v + Kmin]; w ranges between 0 and 1 and can use other operators.

Exponential function

The system builds the lookup table that contains 256 entries by computering:

$$EXPLUT[w] = (\exp(Kmax * w/255.0) - 1) / (\exp(Kmax) - 1)$$

For w = 0,...255. The lookup table values are all ranging from 0 to 1. Given the value v of a pixel ranging between 0 and 1, the system obtains the new grey level value w by w = EXPLUT[255*v]; w ranges between 0 and 1 and we can use other operators.

Figure 5 shows the input image and the image after applying the exponential function with Kmax=4, and the input image and the image after applying the logarithmic function with Kmin=32. The system could generate an infinite number of such images by making varying Kmin and Kmax.

Vertical shadow/ Horizontal shadow

As the two processes are identical, we describe only the vertical shadow process for the abscissas x. This function creates a modification of the grey level of a pixel depending on its spatial position in the image. Here we apply it line by line. An infinite number of functions can be used. For the sake of simplicity, we describe one function. Let X be the width of the image, and let $0<\lambda<1$ real coefficient, and let $m=\lambda*X$. We define the following function f(x), depending on the value of $x \in [0,X]$:

$$F(x) = x / m \text{ if } x \in [0,m]$$

$$F(x) = 1 + (x-m)/(X-m) \text{ if } x \in [m,X].$$

Such a function can be seen in Figure 6. Here X=200 and λ=0.2. For a given abscissa x, F(x) is a coefficient that varies between [0,2].

Given a pixel p=(x,y) with a grey level v, the system obtains in the output image a pixel q=(x,y) with a grey level computed w =v*F(x). Thus, the m first pixels will become darker, the first one at position 0, totally dark, the mth pixel will be unaffected (in Figure 6, the 40th pixel 200*0.2 = 40), and the pixels at abscissa greater than m will become brighter. Figure 7 illustrates the process with 5 different shadow filters, from the left original image.

The horizontal process applies to the ordinate y, and uses the height of the image Y, in exactly the same manner.

2D masks as combination of two 1D masks.

Here we define a mask F such that F(x,y) = Fx(x)*Fy(y). For each pixel p=(x,y) with grey level v, we obtain the new grey level w=v*F(x,y).

2D masks.

These masks are a generalization of the previous mono-dimensional masks, and a coefficient is defined for each pixel F(x,y). The system then obtains a 2D mask that modifies the value v of the pixel p=(x,y), in w = v+F(x,y). These masks can be built manually, by capturing several identical images and by changing the lighting conditions of the room during the image capture. In such a case, if we define the neutral image as I, and all the other images as I_1, I_2, ..., I_N, the system derives N masks by differentiating I with I_1, I_2, ..., I_N: and we obtain F_1 = I-I_1, F_2=I-I_2, ..., F_N=I-I_N. Thus, we modify the value of the pixel p = (x,y), in w=v+F(x,y). Figure 8 illustrates a face capture where the light varies to set up manual lighting masks, by differentiating with the first image. Once again, these masks can be normalized between 0 and 1.

Filter cascading

A synthetic prototype can be obtained by applying several masks to the original normalized image. Let I be the original normalized image, F={F_1,F_2,....F_N} the set of successive masks, and O the obtained synthetic prototype. We have O=F_N(...F_2(F_1(I))...). In some specific cases, we may have F_N(...F_2(F_1(I))...) =

Fn(...F2(F1))(I) = M(I), leading to a much more efficient computation, as the mask $M = F_N(\ldots F_2(F_1)\ldots)$ can be estimated in advance by a combination of masks.

Figure 9 illustrates images obtained during the registration of a user. In this example the preferred system took the 5 leftmost (first column) images to register robustly this person, at a resolution of 15x15 pixels. The grey levels intensity of each image have been re-normalized between 0 and 255.

This set of generated faces could be used in any application using faces as if it was produced online by any camera or video input devices. For example, it can be used in face recognition, face verification, face feature extraction, face retrieval, video conferencing systems, neural network or auto-associative memory training, etc.

By storing multiple possible appearances of the same face, the system increases the probability of retrieving the correct face across lighting conditions, expressions and face directions. The system therefore compensates for the weaknesses of previous methods although does obtain a more complex database.

The system does increase the probability of confusion with another face, but experiments have shown that the best match between two identical persons is statistically higher than the best match between two different persons. As these transforms are applied identically to all the faces, all are penalized or favored in the same way, which is enough to justify this experimental fact. In other words, it is more likely that the registration transforms the face into a future observed face of the same person, than it transforms someone else's face into this future observed face.

The system extends the representations to non-linear geometric transformations and non-linear photometric transformations, to synthesize realistic face prototypes. This is equivalent to an expansion of the training set for

a better generalization of the neural network. The key of this robust registration is to apply allowable transformation to the face image, using face morphing and applying synthesized lighting conditions that approach reality. These transformations are applied to a face for which the eyes coordinates are already known. Another key issue is that the registration process considers errors in the positioning of the eyes that are due to the eyes detection algorithm.

Given a face image and the two eyes positions, the present system is able to synthesize realistic faces from a single face, to provide a robust registration which in turn leads to a robust face recognition/verification algorithm. The consideration of errors in eyes detection, the synthesize of the lighting effects, as well as the face morphing can use pre-registered masks. Their combination allows the creation of a multitude of synthetic faces that are likely to be encountered during the future face recognition tasks.

The subsequent face retrieval algorithm can be straightforward, as the registration process has undertaken most of the difficulties.

The invention has been particularly described with reference to face recognition systems to assist in the understanding of the invention. However, the present invention proposes a way to generate artificial and likely face prototypes from a single face, by creating artificial light effects, artificial face direction and by modeling the localization errors of a face location system, for the purpose of robust face feature extraction or robust face registration. That is the present invention is a facial prototype synthesis system, that generates automatically realistic templates from a single image and not merely a face recognition system. Lighting conditions masks, warping masks and eyes position errors are used to achieve a robust face generation.

Apart from improving face recognition systems the present invention could also be used to improve current face recognition and face verification systems, by transforming and expanding the existing databases containing the faces of the

people, automatically, without recapturing all the photos, under all the face directions, all the lighting conditions, all the scales and rotations.

The present invention therefore proposes a way to generate artificial and likely face prototypes from a single face, by creating artificial light effects, artificial face direction and by modeling the localization errors of a face location system, for the purpose of robust face feature extraction or robust face registration.

Whilst the method and system of the present invention has been summarised and explained by illustrative example it will be appreciated by those skilled in the art that many widely varying embodiments and applications are within the teaching and scope of the present invention, and that the examples presented herein are by way of illustration only and should not be construed as limiting the scope of this invention.

CLAIMS:

1. A face recognition and/or verification system including the step of registering a persons actual face wherein an image of said actual face is captured and synthesized to create a plurality of face prototypes, and wherein said face prototypes are stored for later analysis and comparison with a captured image to be recognised or verified.

2. The system as claimed in claim 1 wherein said face prototypes represent possible appearances of said actual face under various lighting conditions, varying facial expressions, varying facial orientations, and/or modeling errors.

3. The system as claimed in claim 1, wherein comparison of said face prototypes and captured image uses a face matching algorithm.

4. The system as claimed in claim 1, wherein comparison of said face prototypes and captured image uses face templates or feature vectors.

5. The system as claimed in any preceding claim, wherein synthesizing of said actual face includes normalising said actual face image.

6. The system as claimed in claim 5, wherein normalising includes rotating said actual face image to bring eyes of said actual face image to a horizontal plane.

7. The system as claimed in claim 5 or claim 6, wherein normalising includes scaling said actual face image such that the eyes are a fixed distance apart.

8. The system as claimed in claim 7, wherein said eyes are fixed at 50 pixels apart.

9. The system as claimed in any preceding claim wherein the area above the persons eyebrows and below the persons mouth is not synthesized.

10. The system as claimed in any preceding claim wherein synthesizing of said actual face includes determining alternative positions for each eye so as to compensate for possible errors.

11. The system as claimed in claim 10, wherein five alternative positions are determined for each eye.

12. The system as claimed in any preceding claim wherein synthesizing of said actual face includes applying at least one predefined lighting mask to said actual face image.

13. The system as claimed in claim 12, wherein three to 16 predefined lighting masks are used.

14. The system as claimed in any preceding claim wherein synthesizing of said actual face includes applying at least one predefined warping mask to said actual face image.

15. The system as claimed in claim 14, wherein 25 predefined warping masks are used.

16. The system as claimed in claim 12 or claim 13, wherein said at least one lighting mask includes photometric transform.

17. The system as claimed in claim 14 or claim 15, wherein said at least one warping mask includes geometric transform.

18. The system as claimed in claim 17, wherein said geometric transform is estimated using optical flow estimation.

19. The system as claimed in claim 16, wherein said photometric transform includes at least one of:

algorithmic function, exponential stretch, vertical shadow, horizontal shadow and differentiating image.

20. A facial prototype synthesis system wherein an image of a persons actual face is used to create a plurality of face prototypes, said face prototypes representing possible appearances of said actual face under various lighting conditions, varying facial expressions, varying facial orientations, and/or modeling errors, and wherein said face prototypes are stored for later use

21. The system as claimed in claim 20, wherein said actual face image is normalized prior to creating said face prototypes.

22. The system as claimed in claim 21, wherein normalising includes rotating said actual face image to bring eyes of said actual face image to a horizontal plane.

23. The system as claimed in claim 21 or claim 22, wherein normalising includes scaling said actual face image such that the eyes are a fixed distance apart.

24. The system as claimed in claim 23, wherein said eyes are fixed at 50 pixels apart.

25. The system as claimed in any one of claims 20 to 24, wherein the area above the persons eyebrows and below the persons mouth is not synthesized.

26. The system as claimed in any one of claims 20 to 25, wherein to create said face prototypes said system determines alternative positions for each eye so as to compensate for possible errors.

27. The system as claimed in claim 26, wherein five alternative positions are determined for each eye.

28. The system as claimed in any one of claims 20 to 27 wherein to create said face prototypes said system applies at least one predefined lighting mask to said actual face image.

29. The system as claimed in claim 28, wherein three to 16 predefined lighting masks are used.

30. The system as claimed in one of claims 20 to 29 wherein to create said face prototypes said system applies at least one predefined warping mask to said actual face image.

31. The system as claimed in claim 30, wherein 25 predefined warping masks are used.

32. The system as claimed in claim 28 or claim 29, wherein said at least one lighting mask includes photometric transform.

33. The system as claimed in claim 30 or claim 31, wherein said at least one warping mask includes geometric transform.

34. The system as claimed in claim 33, wherein said geometric transform is estimated using optical flow estimation.

35. The system as claimed in claim 32, wherein said photometric transform includes at least one of:

algorithmic function, exponential stretch, vertical shadow, horizontal shadow and differentiating image.

36. The system as claimed in claim 20, wherein said face prototypes are generated by applying photometric and/or geometric transforms to said image.

37. A facial prototype synthesis system wherein an image of a persons actual face is normalised and synthesized by determining possible alternative eye positions and applying at least one mask to said image to create a plurality of face prototypes, and wherein said face prototypes represent possible appearances of said actual face under various lighting conditions, varying facial expressions, varying facial orientations, and/or modeling errors.

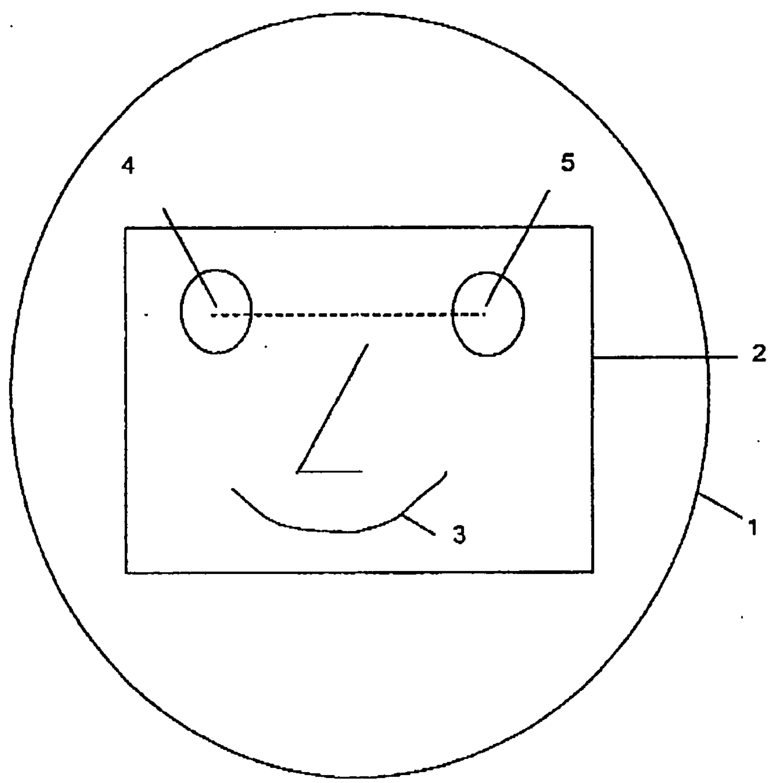


Figure 1

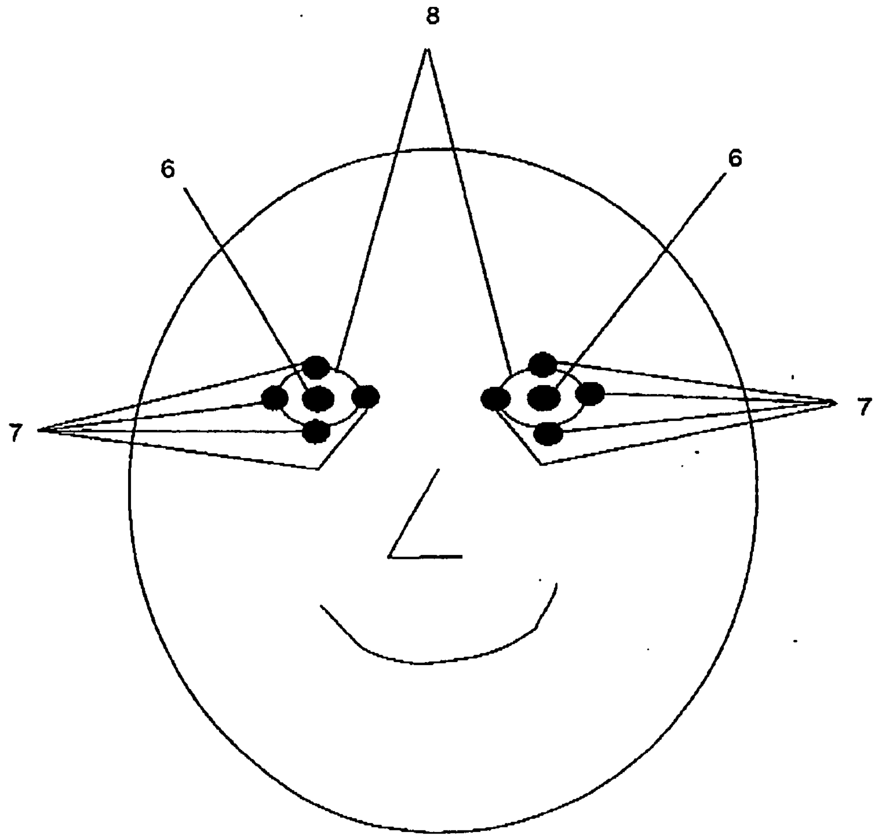


Figure 2

Figure 3

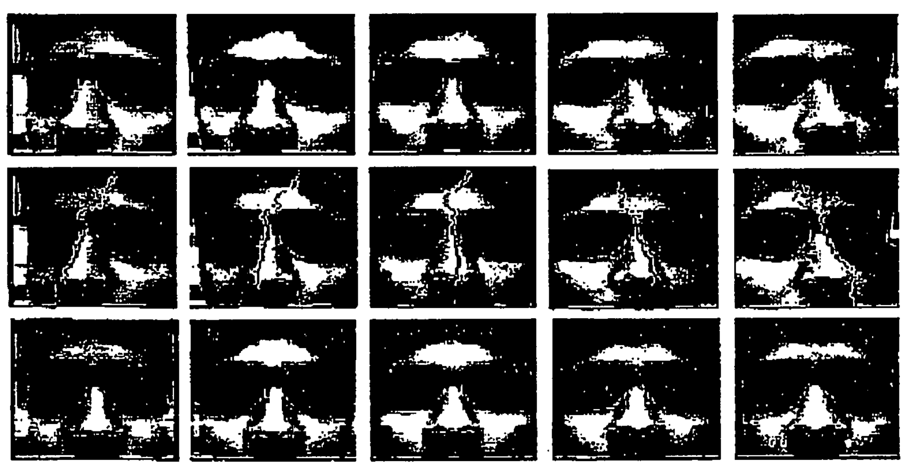


Figure 4

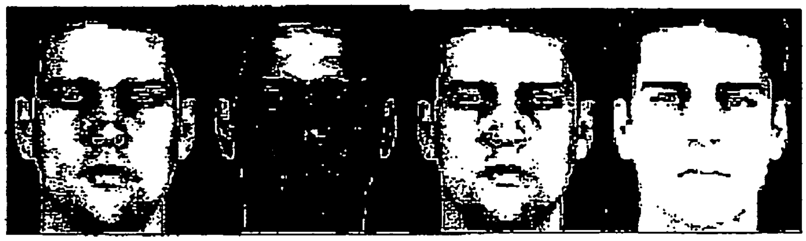


Figure 5

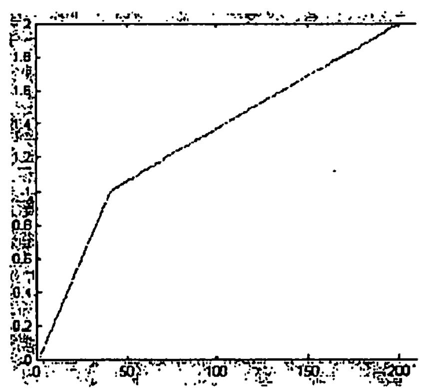


Figure 6

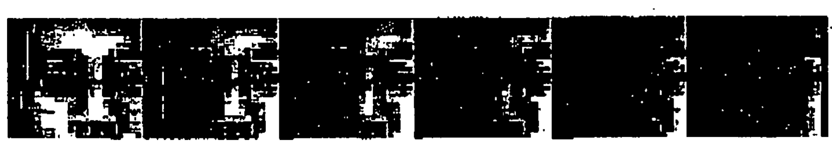


Figure 7

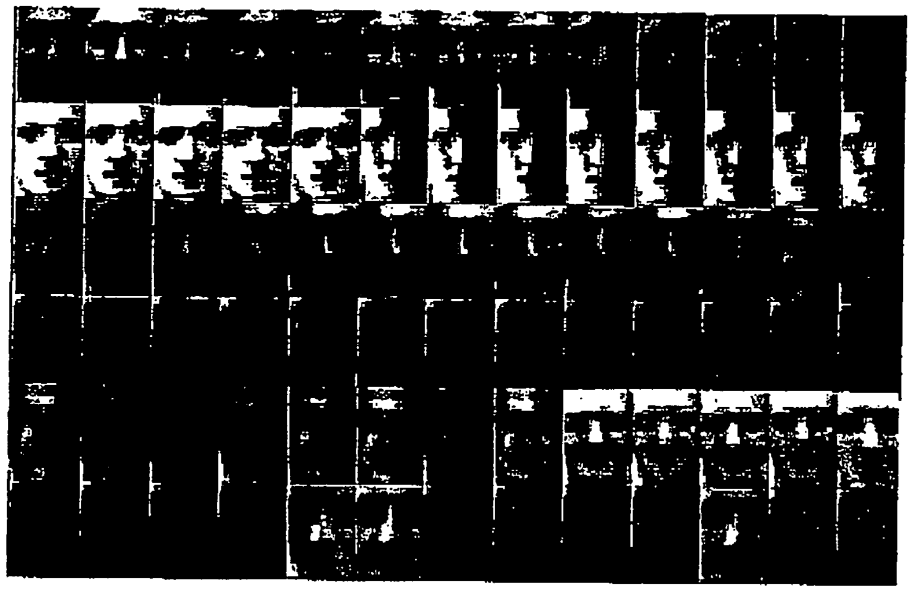


Figure 8

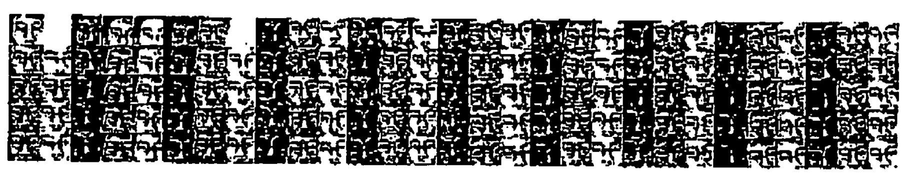


Figure 9